Proceedings of the Eleventh Biennial Symposium on Arts and Technology,
February 28-29, 2008, New London, Connecticut, pp. 174-179

A System for Searching Sound Palettes

Stephen V. Rice
The University of Mississippi
P.O. Box 1848, University, MS 38677
(662) 915-5359; rice@cs.olemiss.edu

Stephen M. Bailey
Comparisonics Corp.
P.O. Box 1960, Grass Valley, CA 95945
(530) 272-0811; sbailey@comparisonics.com

ABSTRACT

Like painters selecting hues from color palettes, sound artists and composers choose from palettes of available sounds. Their choices become part of plays, films, radio and television programs, songs, animations, and games. *FindSounds Palette* is an audio retrieval system designed to facilitate this creative process. Local audio files, and audio files on the Web, are searched by text description and sonic similarity. Sounds produced by changing the speed of audio recordings can also be searched, which multiplies the size of the palettes. Waveform displays, colored to represent frequency information, facilitate audio editing and serve as visual thumbnails in query results.

INTRODUCTION

In the field of music information retrieval, techniques have been proposed for searching song collections based on melody and musical style [3]. In spoken document retrieval, speech recordings are converted to phonetic or text representations and these transcripts are searched [2]. These forms of audio retrieval have received more attention in the research community than searching collections of sound effects and musical instrument samples. However, there is a centuries-old need to locate sounds for creative use in the arts.

Sound effects were used in the ancient Greek theatre of Aeschylus, Euripedes, and Sophocles [11]. In Elizabethan theatre, scripts called for the sounds of alarms, chimes, and gunshots [8]. In the 1930s, the production of sound effects for film, theatre, and radio increased in sophistication. Thousands of prerecorded sounds became available on 78 rpm phonograph records [1], and "manual" sound effects were created by clever use of an enormous variety of objects and devices. An "alphabetical glossary" at NBC Radio contained thousands of techniques for sound generation [4]. Such a cookbook might include the following recipes [8,18]:

- *Fire*—crinkle cellophane; the faster you crinkle, the bigger the fire;
- *Rain*—sprinkle salt on paper;
- *Walking in mud*—handle a soggy newspaper.

An inertia starter for an old prop plane created the sound of the Tasmanian Devil spinning wildly in Warner Brothers cartoons [9]. The ghost sounds in the 1989 movie, *Ghostbusters II*, were produced by a rice steamer [9]. Today's sound designers for film and television continue to rely on gadgets for sound production and also utilize digitized collections of 100,000 or more sound-effect recordings.

In the early 20th century, composers such as Bartók, Debussy, Strauss, and Stravinsky expanded the percussion section of orchestras and devised novel techniques for playing traditional instruments to obtain new sounds [5,16]. Satie's *Parade*, incorporating sirens, starting pistols, typewriter, and foghorn, caused a scandal when performed in Paris in 1917; conservative listeners considered it blasphemous for music to include such sounds [15]. Pierre Schaeffer's pioneering 1948 composition, *Étude aux Chemins de Fer*, is a fascinating montage of sounds recorded at the Paris train depot and demonstrated that any sound is raw material for creative use [15].

The arrival of electronic sound synthesizers was heralded by many composers. Edgard Varèse extolled the "electronic medium" for adding "an unbelievable variety of new timbres to our musical store," and for "the possibility of obtaining any differentiation of timbre, of sound-combinations, and new dynamics far beyond the present human-powered orchestra" [15]. The Moog synthesizer of the 1960s was the first synthesizer to be mass produced. Today there are innumerable hardware and software synthesizers available. In addition, many composers utilize digital samples from collections containing 50,000 or more natural and synthesized notes, chords, and drum beats.

While modern technology has enabled sound designers and composers to amass large palettes of digitized sounds, their ability to effectively search these collections has been limited. Creative people seek the best access to the most sounds. The challenge for computer scientists is to develop ways to increase the accessibility and size of sound palettes. With these goals in mind, the authors of this paper developed a unique audio retrieval system named *FindSounds Palette*. The first version of this system was introduced in 2002. Current users of this system include sound designers, musicians, filmmakers, animators, and game developers.

METADATA SEARCH

A collection of audio files stored on local hard drives is indexed by *FindSounds Palette* and is called MyPalette.

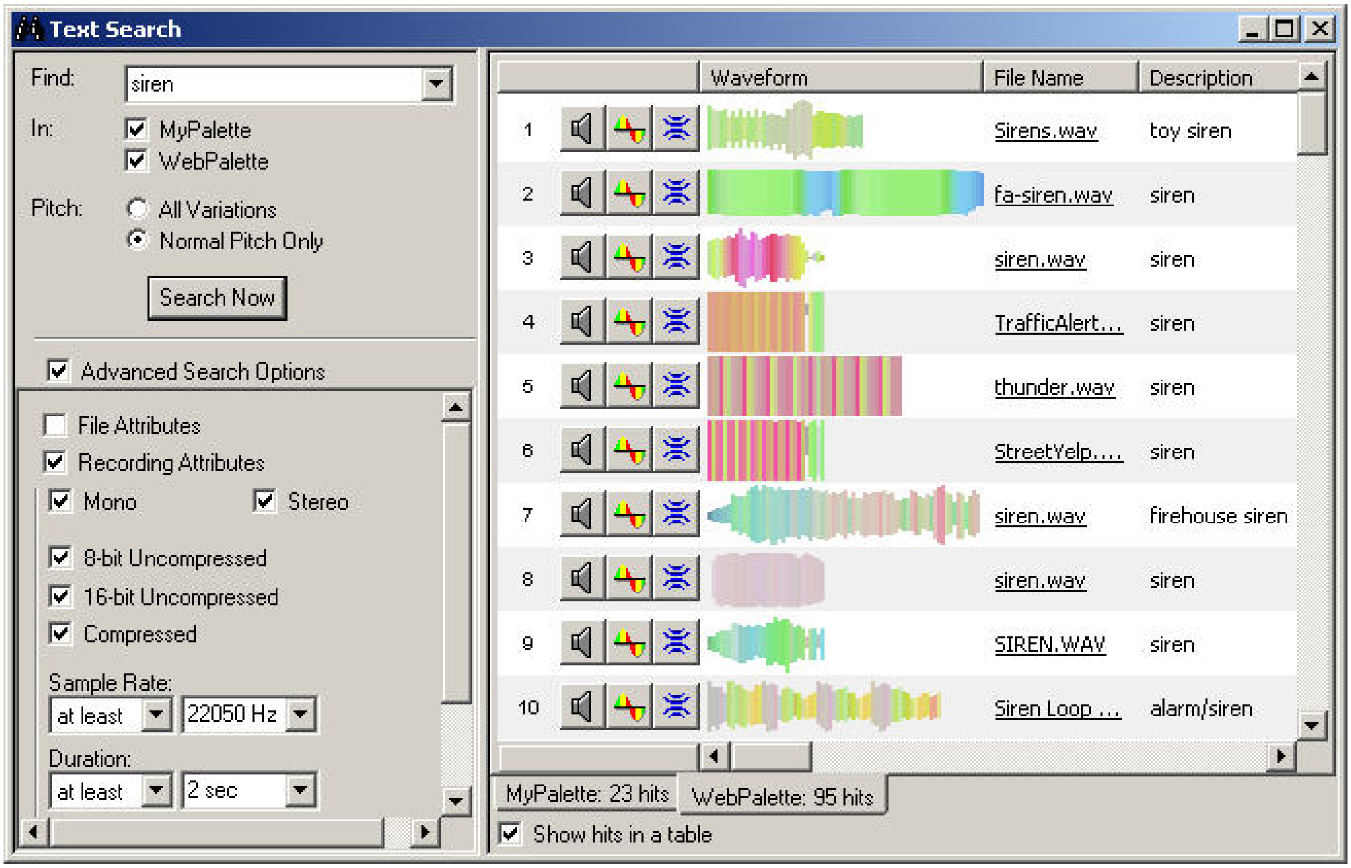
Tens of thousands of audio files may be cataloged. All available metadata is potentially valuable for retrieval purposes. *FindSounds Palette* extracts some metadata automatically from audio files, such as file name, file format, file size, number of channels (mono or stereo), sample rate, bit resolution, file compression, and duration. The user may supply additional information in free-form text fields named Description, Source, Copyright, Notes, and Genre. Each audio file can be placed in a class (Effect, Instrument, or Other) and in any user-defined category and sub-category. This hierarchical organization is easily manipulated. Additional metadata fields intended for musical instrument samples include key and tempo (in beats per minute).

A text search of MyPalette is launched by entering one or more words in a search box. Queries may be qualified using any combination of search criteria. Minimum and maximum values may be specified for file size, sample rate, duration, and tempo. Specific file formats, number of channels, and musical key may be requested.

Search results may be displayed in a table, with one row for each hit and one column for each metadata field, or the user may select the "stacked" view in which the metadata for each hit is formatted on consecutive lines. In the tabular view, the hits may be sorted by clicking on column headings.

Figure 1 shows the results of a text search for "siren." The user clicks on the Play icon next to a hit to play the audio file or clicks on the Open icon to open the audio file in an audio editor window.

Notably, each hit has a colored waveform display showing the first ten seconds of the audio recording. This display is a graph of amplitude versus time which has been color-coded to convey the frequency content of the audio signal. Sounds dominated by high frequencies receive more red component and mid-range sounds are colored by shades of green or blue. Low (bass) sounds are assigned dark colors, and noisy sounds, such as white noise, appear in grey. Similar sounds are represented by similar colors, and changes in sound are made evident by changes in color. This display serves as a "visual thumbnail." By inspecting these thumbnails, the user may decide which hits to audition. This display has proven to be immensely helpful. For more information, see references [12] and [14].



Find: siren

In: ☑ MyPalette ☑ WebPalette

Pitch: ○ All Variations ● Normal Pitch Only

☑ Advanced Search Options

□ File Attributes

☑ Recording Attributes

☑ Mono ☑ Stereo

☑ 8-bit Uncompressed

☑ 16-bit Uncompressed

☑ Compressed

Sample Rate: at least 22050 Hz

Duration: at least 2 sec

	Waveform	File Name	Description
1		Sirens.wav	toy siren
2		fa-siren.wav	siren
3		siren.wav	siren
4		TrafficAlert…	siren
5		thunder.wav	siren
6		StreetYelp.…	siren
7		siren.wav	firehouse siren
8		siren.wav	siren
9		SIREN.WAV	siren
10		Siren Loop …	alarm/siren

MyPalette: 23 hits WebPalette: 95 hits

☑ Show hits in a table

Figure 1. Results of a text search for "siren" using *FindSounds Palette*.

SOUNDS-LIKE SEARCH

Onomatopoeia is the formation of words to imitate sounds, for example, buzz, crunch, hiss, pop, screech, and thud. People who catalog sounds have raised onomatopoeia to an art form in desperate attempts to describe sounds. The following descriptions appear in a current sound-effects catalog: gedunk, kablam, kabong, pingy wobbles, wiggle bowang. Catalogers work overtime to find the right adjectives: "searing harmonic slashes," "industrial amorphous textured presence," "incendiary fuzz mutations." According to Hollywood sound editor Mark Mangini, sound editors have invented thousands of words to describe sounds, such as boink, boing, twang, twung, squidge, zip, rico, whibble, wobble, and wubba, yet every sound editor uses a different set of terms [9]. Such descriptions convey little information, do not translate well to other languages, and are nearly useless for keyword searches.

Describing the source of a sound, if known, is often much easier than describing the sound itself, and most catalogers resort to this approach. Most of us know the sounds of an "automobile idling," "several coins dropped on a tile floor," and a "roller coaster passing by." Source descriptions are less useful if we are unfamiliar with the sounds, for example, "llama vocalizing," "slab of steel emerging from a furnace," and "water lock gates opening." If the source of a sound is a synthesizer, then how should it be described? Consider a synthesized sound used in a *Star Trek* movie to warn that the dylithium crystals are going to overload [9]. "Weird electronic sound" and "dylithium crystal alarm" are clearly inadequate for retrieval purposes. A synthesizer can generate thousands of sounds that cannot meaningfully be expressed in words.

The source of a sound is of little interest to a sound designer who intends to use the sound for something else. In fact, knowing the source makes it harder to evaluate the sound. It is difficult to imagine that a cat can create the sound of a monster, but if you don't know that a sound came from a cat, you can listen to it objectively. In his book on sound effects, Robert L. Mott encourages sound designers to "disassociate the names of the sounds with the sounds themselves" and to "concentrate on the sound" and "ignore its source" [11]. *Star Wars* sound designer Ben Burtt makes it a practice to play sounds for the director without revealing their source so that the director will listen to them uninfluenced by their origin [9]. *Jurassic Park* sound designer Gary Rydstrom believes the most important talent for sound design is the ability to separate what a sound *is* from how it is made [18].

The "semantic gap" in image retrieval is well known [17]. Indeed, without text labels, it is as difficult for a computer to find the sounds of dogs as it is for a computer to locate pictures of dogs. The added difficulty of meaningfully describing sounds in words suggests that the semantic gap is a prohibitive obstacle to sound retrieval. Yet the interest of creative people in the sounds themselves, and their disinterest in text descriptions of sounds, provides an opportunity to circumvent the semantic gap. In a "sounds-like search" or "query by sound example," any example sound is presented to a computer, and the computer finds those sounds in the collection that are most perceptually similar to the example sound, regardless of how or if the sounds are described in words. This is a form of "content-based" audio retrieval.

S. V. Rice, an author of this paper, developed one of the first "sound-matching" algorithms and it is incorporated into *FindSounds Palette*. Any sound can serve as the example sound in a sounds-like search, and *FindSounds Palette* responds by retrieving the most similar sounds. A similarity score, on a scale from zero (least similar) to 100 (most similar, i.e., identical), is shown for each hit, and the hits are displayed in decreasing order of their similarity scores so that the best matches appear first.

The sound-matching algorithm estimates the similarity of sounds even if they differ in duration, format, sample rate, compression, and number of channels. The matches are determined based on perceptual characteristics extracted by the algorithm from the audio signal and are uninfluenced by text descriptions. As a result, the sound of a revving engine may match a growling tiger, screeching tires may match a ranting chimpanzee, and a tympani roll may match a rumble of thunder. Such matches are of interest to sound designers but would never be discovered from text descriptions. The example sound may be a hit returned by a text search or by a sounds-like search. After a sounds-like search yields an interesting match, that match can be used as the example sound in a subsequent sounds-like search. This process can be repeated to explore the collection based on sonic similarity.

A sounds-like search may be qualified by any metadata criteria. For example, the user may request that only those sounds labeled "engine" be retrieved and ranked by similarity to an example engine sound. Creatively applied, this capability can be used to find coyote howls that sound like a particular siren and to locate saxophone samples that resemble an elephant's bellow.

SEARCHING MULTIPLE SPEEDS

Phonographs with variable speed control were needed in the 1920s to play "78 rpm" records because the speed at which they were actually recorded ranged from 70 to 85 rpm. Interesting sounds can be created by slowing down or speeding up a recording, and the speed control became a valuable tool of the sound designer [1]. Hindemith and Toch composed short pieces using phonographic speed change [10], and Schaeffer and Varèse experimented considerably with the technique. Mott played a single recording of a waterfall at different speeds to create the sounds of ocean surf, city traffic, a jet airplane, an atomic bomb explosion, and a printing press [11]. In the 1989 movie, *Indiana Jones and the Last Crusade*, a recording of chickens was speeded up and used as the sound of a cave filled with rats [6]. Walter Murch, regarded as the dean of sound designers, would change the speed of a sound (e.g., the outboard motor in the 1974 movie, *Godfather II*) so that it would harmonize with the background music and prevent dissonance [7].

FindSounds Palette introduced the technique of searching a collection of sounds at multiple speeds. Each audio file may be indexed at the normal speed and 24 additional speeds: the normal speed increased by one to 12 semitones, and the normal speed decreased by one to 12 semitones. The user may specify whether only normal-speed hits are to be retrieved, or if all speed variations are to be searched. For the latter, the speed of each hit is displayed as either normal or some semitone increment or decrement. Each audio file is stored only once on disk at its normal speed. When a speed variation is auditioned, the file is read from disk and played at the requested speed.

The consequence of this capability is that a database of 10,000 audio files becomes a searchable collection of 250,000 sounds. When the user auditions a speed variation, it may be the first time it is heard by human ears. The speed variations are impossible to describe in words but are made accessible by a sounds-like search.

WEB SEARCH

FindSounds.com is the first Web search engine for sound effects and musical instrument samples [13]. It was launched in August 2000 and was developed by the authors of this paper. It incorporates the sounds-like search and was the first Web search engine with content-based audio retrieval. It currently processes each month more than 2 million queries for more than 250,000 users.

FindSounds.com locates audio files on the Web. Requests from early FindSounds.com users for a "home version" led to the development of *FindSounds Palette* for searching local sound databases. However, in addition to local search, *FindSounds Palette* provides greater access to Web audio files than FindSounds.com.

In *FindSounds Palette*, the collection of indexed audio files on the Web is called WebPalette. The target of any search may be MyPalette, WebPalette, or both. If both are specified, two lists of hits are returned, one containing local hits from MyPalette and the other containing remote hits from WebPalette. Clicking on a remote hit causes an audio file to be downloaded to the user's computer and played. Remote hits can be saved locally in MyPalette.

All of the search features we have described can be used for WebPalette searches. The index of Web audio files is maintained at the FindSounds.com server. *FindSounds Palette* sends a WebPalette query to the server which responds by returning a list of hits. There are currently about 50,000 indexed Web audio files. However, each audio file is indexed at approximately 40 speeds. Therefore, a sounds-like search of WebPalette searches approximately 2 million sounds.

AUDIO RECORDING AND EDITING

People working with sounds need the ability to record and edit sounds. Thus, *FindSounds Palette* has an integrated audio recorder and editor. The audio file being recorded or edited is displayed as a colored waveform which the user may pan and zoom. The coloring helps the user to see the sounds which greatly facilitates the editing process. Editing operations include cut, copy, paste, mix, trim, delete, adjust volume, fade in/out, undo, and redo. The speed of the file may be changed and the waveform is automatically recolored to represent the modified frequency content.

The ability to search what you are editing is commonplace in word processing but not in audio editing. Searching within an audio recording is typically limited to finding amplitude spikes or text-annotated sections of the recording. In the *FindSounds Palette* audio editor, however, the user may select any sound in the waveform display and the program automatically highlights similar sounds in the display. The user may advance the cursor forward or backward in the recording from one match to the next. For example, the user can select an example bass drum beat and then automatically find each matching bass drum beat. This ability to navigate within an audio recording based on sonic similarity is described in reference [14]. The user may adjust the "matching threshold" to specify the minimum similarity score of a match. Lowering the threshold identifies more matches, and raising the threshold marks fewer matches.

In addition to searching within an audio file, any sound may be selected in the waveform display and used as the example sound in a sounds-like search of MyPalette and WebPalette. The sound may be from an impromptu recording made using the audio recorder in which the user mimics a desired sound into a microphone using his or her voice or using props. The recording can be edited or speed-changed to "fine tune" it before launching a search for similar sounds.

Figure 2a shows the colored waveform display of a recording of a whale that has been speeded up by four semitones. The first part of the recording has been selected by the user (indicated by the black background) and is the example sound in a sounds-like search of WebPalette. Figure 2b presents a list of hits in order of decreasing similarity scores. Because the hits sound similar to the example, their waveforms have similar colors. Each hit is a speed variation indicated by a positive or negative number of semitones. The example sound matched speed-altered recordings of whales, loons, a sparrow, a mosquito, human burps, radar beeps, a bell, a whimpering gorilla, a screaming toad, a Japanese wood flute, radio beacons, and the routing tone used by the Irish telephone system.

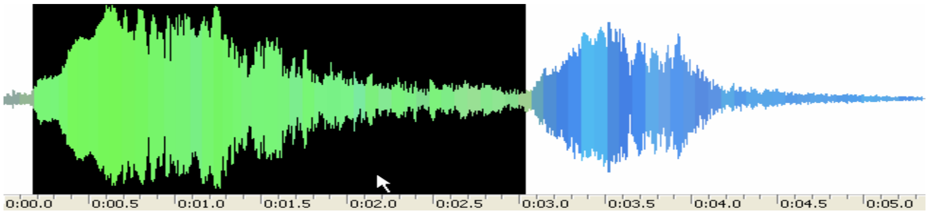
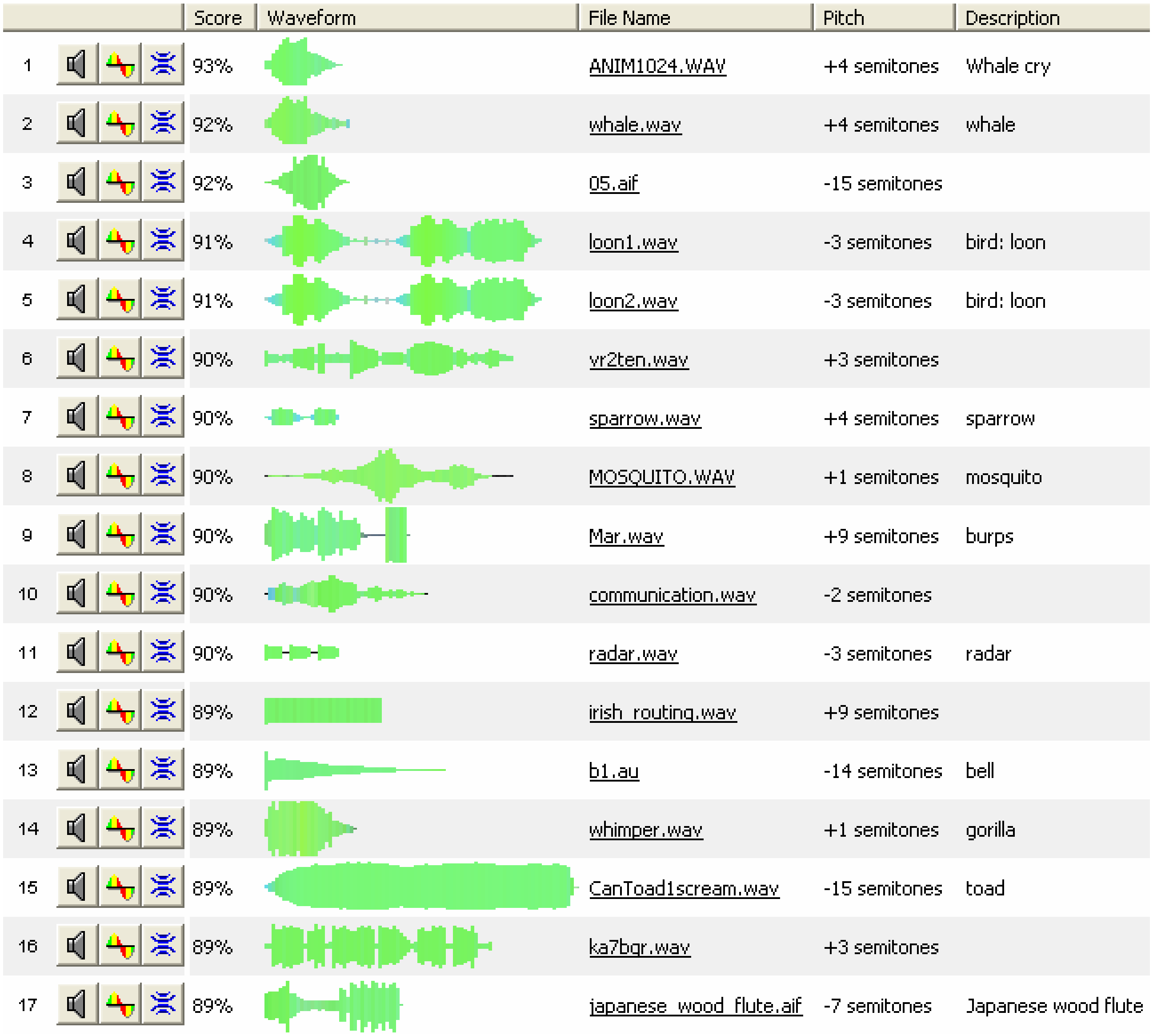


Figure 2a. Selecting an example sound for a sounds-like search.



		Score	Waveform	File Name	Pitch	Description
1		93%		ANIM1024.WAV	+4 semitones	Whale cry
2		92%		whale.wav	+4 semitones	whale
3		92%		05.aif	-15 semitones	
4		91%		loon1.wav	-3 semitones	bird: loon
5		91%		loon2.wav	-3 semitones	bird: loon
6		90%		vr2ten.wav	+3 semitones	
7		90%		sparrow.wav	+4 semitones	sparrow
8		90%		MOSQUITO.WAV	+1 semitones	mosquito
9		90%		Mar.wav	+9 semitones	burps
10		90%		communication.wav	-2 semitones	
11		90%		radar.wav	-3 semitones	radar
12		89%		irish_routing.wav	+9 semitones	
13		89%		b1.au	-14 semitones	bell
14		89%		whimper.wav	+1 semitones	gorilla
15		89%		CanToad1scream.wav	-15 semitones	toad
16		89%		ka7bgr.wav	+3 semitones	
17		89%		japanese_wood_flute.aif	-7 semitones	Japanese wood flute

Figure 2b. Results of a sounds-like search.

FUTURE DIRECTIONS

In addition to MyPalette and WebPalette, additional sound palettes can be made accessible from *FindSounds Palette*, such as commercial libraries of sound effects and musical instrument samples. The authors developed a prototype PeerPalette that permits *FindSounds Palette* users to share their MyPalette collections with one another via a peer-to-peer mechanism. This capability can be added to the system pending resolution of copyright issues in peer-to-peer networks.

In sound production for movies, sound-making devices are used on "Foley stages" to embellish the movie soundtrack with sound effects [18]. The sounds produced by these gadgets can be recorded and searched by sound similarity. A retrieved sound can be accompanied by the recipe for producing it.

Today's synthesizers can produce more sounds than can be heard in a lifetime. Users must explore the sounds through the tedious process of setting the synthesis parameters, playing a sound, changing the parameters, playing another sound, changing the parameters again, and so on. Wouldn't it be wonderful to perform a sounds-like search of the universe of sounds that can be produced by a synthesizer? The user could audition the sounds in a list of hits, and for each sound, obtain the parameter settings used to generate it. The computer music research community has for decades focused on synthesizing new sounds. We believe that it is time to focus on ways to search the nearly infinite variety of available sounds.

Lastly, we note that the technique demonstrated by *FindSounds Palette* for searching speed variations at semitone intervals is only the beginning of what can be done to enlarge sound palettes. More speed variations can be searched and at finer intervals. (As Ferruccio Busoni exclaimed in his 1907 *Sketch of a New Aesthetic of Music*, "Nature created an infinite gradation—infinite!" [15]). Sounds can be transformed by many techniques including filtering, reverberation, modulation, chorusing, flanging, and phasing. Sounds can be overlaid and juxtaposed in unlimited ways. The challenge for computer scientists is to provide ways to search the combinatorial explosion of sounds derivable from existing palettes.

REFERENCES

1. Brunelle, R. The art of sound effects, part 2. *Experimental Musical Instruments, 12,* 2 (1996), 68-74.
2. Coden, A., Brown, E.W., and Srinivasan, S., (eds.). *Information Retrieval Techniques for Speech Applications, LNCS 2273*. Springer-Verlag, 2002.
3. Downie, J.S. Music information retrieval. *Annual Review of Information Science and Technology, 37,* Cronin, B., (ed.), Information Today, Medford NJ, 2003, 295-340.
4. Educational Radio Script Exchange. *Handbook of Sound Effects*. U.S. Office of Education, Washington DC, 1940.
5. Holland, J., and Page, J.K. Percussion. *Grove Music Online*, Macy, L., (ed.), Oxford University Press, 2004. Available at http://www.grovemusic.com.
6. Holman, T. *Sound for Film and Television*. Focal Press, Boston MA, 1997.
7. Jarrett, M. Sound doctrine: An interview with Walter Murch. *Film Quarterly, 53,* 3 (2000), 2-11.
8. Kaye, D., and LeBrecht, J. *Sound and Music for the Theatre: The Art and Technique of Design*. Focal Press, Boston MA, 2000.
9. LoBrutto, V. *Sound-on-Film: Interviews with Creators of Film Sound*. Praeger, Westport CT, 1994.
10. Luening, O. Some random remarks on electronic music. *J. Music Theory, 8,* 1 (1964), 89-98.
11. Mott, R.L. *Sound Effects: Radio, TV, and Film*. Focal Press, Boston MA, 1990.
12. Rice, S.V. Frequency-based coloring of the waveform display to facilitate audio editing and retrieval, in *Proceedings of the 119th Convention of the Audio Engineering Society* (New York NY, October 2005), AES, New York NY, 2005, paper #6530.
13. Rice, S.V., and Bailey, S.M. A web search engine for sound effects, in *Proceedings of the 119th Convention of the Audio Engineering Society* (New York NY, October 2005), AES, New York NY, 2005, paper #6622.
14. Rice, S.V., and Patten, M.D. *Waveform Display Utilizing Frequency-Based Coloring and Navigation*. U.S. patent 6,184,898, Patent and Trademark Office, Washington DC, 2001.
15. Russcol, H. *The Liberation of Sound: An Introduction to Electronic Music*. Prentice-Hall, Englewood Cliffs NJ, 1972.
16. Simms, B.R. *Music of the 20th Century: Style and Structure*. Schirmer, Woodbridge CT, 1996.
17. Smeulders, A.W.M., et al. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Analysis and Machine Intelligence, 22,* 12 (2000), 1349-1380.
18. Sonnenschein, D. *Sound Design: The Expressive Power of Music, Voice, and Sound Effects in Cinema*. Michael Wiese Productions, Studio City CA, 2001.